A(r)Ray of Hope in Analysis of the Function and Diversity of Microbial Communities

MARTIN F. POLZ*, STEFAN BERTILSSON†, SILVIA G. ACINAS, AND DANA HUNT

*Department of Civil and Environmental Engineering, Massachusetts Institute
of Technology, Cambridge, Massachusetts 02139*

The vast majority of microorganisms in the environment remain uncultured, and their existence is known only from sequences retrieved by PCR. As a consequence, our understanding of the ecological function of dominant microbial populations in the environment is limited. We will review microbial diversity studies and show that these may have moved from an extreme underestimation to a potentially severe overestimation of diversity. The latter results from a simple PCR-generated artifact: the cloning of heteroduplex molecules followed by Escherichia coli *mismatch repair, which may generate an exponential increase in observed sequence diversity. However, simple modifications to current PCR amplification protocols minimize such artifactual sequences and may bring within our reach estimation of bacterial diversity in environmental samples. Such estimates may spur new culture-independent approaches based on genomic and microarray technology, allowing correlation of phylogenetic identity with the ecological function of unculturable organisms. In particular, we are developing a DNA microarray that enables identification of individual populations active in utilization of specific organic substrates. The array consists of 16S and 23S rDNA-targeted oligonucleotides and is hybridized to RNA extracted from samples incubated with ^{14}C-labeled organic substrates. Populations that metabolize the substrate can be identified by the radiolabel incorporated in their rRNA after only one*

* To whom correspondence should be addressed. E-mail: mpolz@mit.edu

†Present address: Department of Evolutionary Biology, Limnology, Uppsala University, Norbyvägen 20, SE-75236, Uppsala, Sweden.

to two cell doublings, ensuring realistic preservation of community structure. Thus, the microarray approach may provide a powerful means to link microbial community structure with in situ function of individual populations.

The last two decades have seen a radical shift in our understanding of microbial diversity. Previously the number of bacterial and archaeal species had been estimated to be in the thousands. It is now generally accepted that the number may actually be as high as several million (Torsvik *et al.*, 2002). This change was brought about by a gradual replacement of diversity estimates based on pure-culture isolation of strains with a determination of diversity based on co-occurring gene sequences, largely ribosomal RNA (rRNA) genes (Head *et al.*, 1998). Although early attempts were made to screen diversity by shotgun cloning of environmental DNA, with subsequent detection and sequencing of rRNA gene inserts (Schmidt *et al.*, 1991), large-scale application of the molecular approach was dependent on PCR protocols that allow the enrichment of rRNA genes from genome mixtures using universal primers (Head *et al.*, 1998). Today, the assessment of the entire diversity of rRNA sequences (ribotypes) coexisting within specific microbial communities has become a realistic possibility due to the ease of PCR implementation and the increased availability of high-throughput sequencing facilities.

Although the exact magnitude of microbial diversity still remains an open question, the PCR-based approach has led to the retrieval of large numbers of sequences from almost any environment examined (Hugenholtz *et al.*, 1998). Thus extensive comparative databases are now available from which patterns of microbial community structure are beginning to emerge. For example, studies of bacterioplankton diversity in the ocean, which represents one of the best-studied environments, have shown that, surprisingly, the major phylogenetic groups in the open ocean and the coastal ocean are similar, despite marked differences in trophic

state and habitat quality between these environments (Giovannoni and Rappé, 2000). Such observations have provided important insights; yet they also highlight the major problem of the molecular diversity approach: because very few of the retrieved sequences have closely related cultured representatives available, the ecological role of an organism in question cannot even be guessed (Hugenholtz et al., 1998). Furthermore, even when closely related cultured organisms exist, they can display quite significant genomic, physiological, or metabolic differences (Gray and Head, 2001). New alternatives for diversity studies, such as analysis of large genome fragments retrieved from the environment (Béja et al., 2000) or gene cassette PCR for recovery of complete open reading frames from environmental DNA (Stockes et al., 2001), can enhance our understanding of uncultured organisms. Nonetheless, elucidation of structure-function relationships or niche differentiation of populations within microbial communities remains one of the big challenges in microbial ecology.

During the last few years, molecular diversity studies have been augmented with tracer techniques that allow assignment of biogeochemical function to uncultured microbial populations [recently reviewed by Gray and Head (2001)]. Most notably, combined microautoradiography and in situ hybridization (STAR- or MICROFISH) (Lee et al., 1999; Ouverney and Fuhrman, 1999; Cottrell and Kirchman, 2000) or stable isotope probing (Boschker et al., 1998; Radajewski et al., 2000) allow identification of microbial populations responsible for the metabolism of specific organic compounds. In both cases, environmental samples are incubated with isotopically labeled substrates. In STAR- or MICROFISH, microautoradiography and in situ hybridization are carried out on the same microscope slide with the goal of matching uptake of radiochemicals with phylogenetic identification on the single-cell level. In stable isotope probing, either lipid biomarkers (Boschker et al., 1998) or DNA (Radajewski et al., 2000) are extracted from communities incubated with ^{13}C-labeled compounds. If cells grow on the added compounds, their pool of macromolecules will be isotopically heavy compared to those of metabolically inactive organisms. This makes it possible to identify the organism in one of two ways: (1) by mass spectrometry of labeled "signature" lipids (Boschker et al., 1998); or (2) by separation by ultracentrifugation of community DNA according to mass differences, followed by identification of rRNA genes in the isotopically heavy DNA pool by PCR, cloning, and sequencing (Radajewski et al., 2000).

The above approaches have already produced interesting insights into the ecological roles of uncultured Bacteria and Archaea. For example, using MICROFISH, it was demonstrated that low-temperature Archaea, which represent a dominant group in deep ocean water but are currently known only from rRNA gene clone libraries, readily take up amino acids at low ambient concentrations (Ouverney and Fuhrman, 1999). In another study, incubation of anaerobic sediments with ^{13}C-acetate yielded signature lipids of gram-positive bacteria rather than those of the more readily isolated delta Proteobacteria sulfate-reducing bacteria (Boschker et al., 1998). However, each of the techniques has distinct drawbacks. Stable isotope probing requires very high substrate concentrations and long incubation times, so that the procedure actually resembles enrichment cultures (Radajewski et al., 2000); MICROFISH involves labor-intensive hybridization and microautoradiography, which limits the number of populations whose metabolism can be explored.

We are currently developing a combination of DNA microarrays and radiotracer incubations, the "functional diversity array," as a high-throughput complement to the above methods (Bertilsson and Polz, 2001). DNA microarrays can carry hundreds to thousands of specific nucleic acid probes, which are arrayed in discrete spots. During the hybridization process, these probes capture their specific target from mixtures of templates. If these are either radioactively or fluorescently labeled, the presence or, to a certain extent, the quantity of all specific templates for which probes have been spotted can be ascertained. The application of DNA microarrays for screening and monitoring microbial community structure by arraying rRNA-specific oligonucleotide probes is being explored by a number of laboratories (Cho and Tiedje, 2001; Small et al., 2001; Koizumi et al., 2002). In the functional diversity array, diversity screening is combined with detection of populations responsible for specific transformations in the community (Fig. 1). Samples are spiked with ^{14}C-labeled compounds, leading to incorporation of radionuclides into rRNA of populations that actively metabolize the compound of interest (Bertilsson and Polz, 2001). RNA is subsequently extracted, fluorescently labeled, and hybridized to the microarray, which contains oligonucleotide probes specific for each "ribotype" in the community. Radioactivity in each spot due to hybridization can be determined by either microautoradiography or phosphor-imaging, so that in combination with the fluorescent signal, a specific activity can be estimated for each population (Fig. 1).

For the functional diversity array to be generally applicable, differentiation of populations by the arrayed probes is not, by itself, sufficient. In addition, several critical questions must be evaluated. First, what is the detection limit for ^{14}C-labeled rRNA hybridized to the array? Second, can realistic substrate concentrations be used, and what are the kinetics of rRNA synthesis after uptake of label under environmental conditions? Third, to what extent can an entire microbial community be represented on the array? Below, we evaluate these questions, with special emphasis on approaches for studying rRNA gene diversity in microbial communities as a necessary precondition for determining biogeochemical activity of previously unidentified populations.

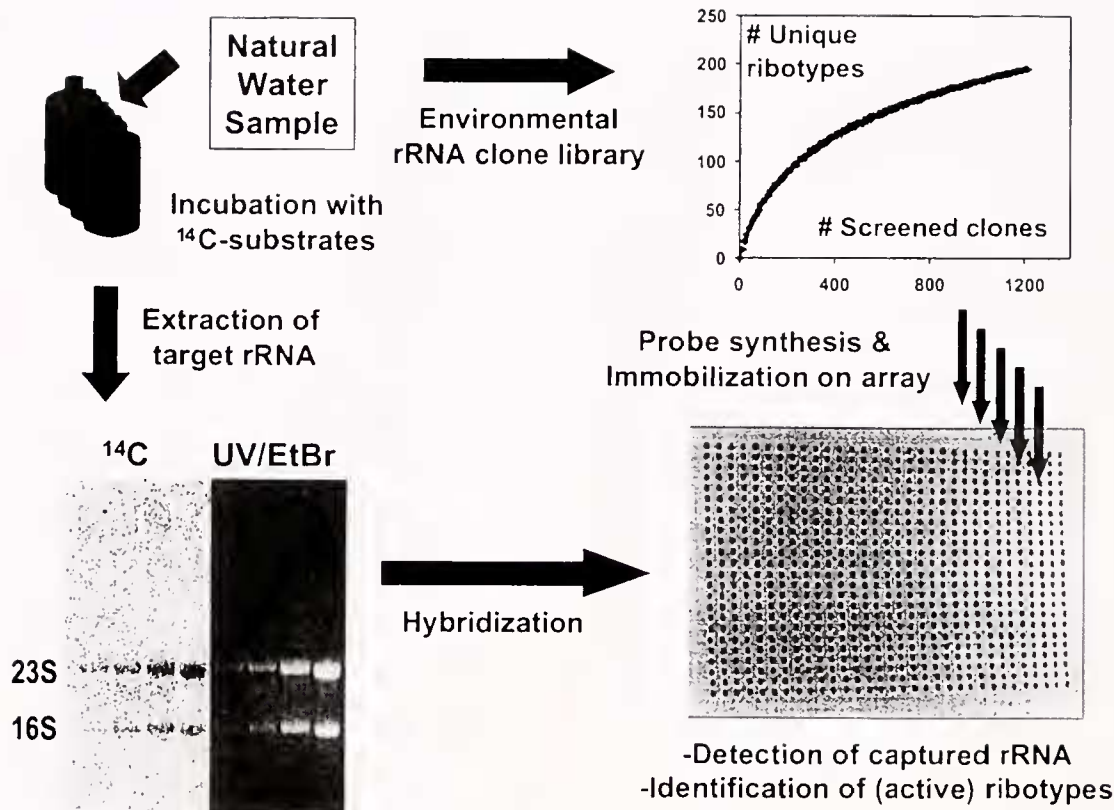Quantification of the radioactive signal on arrays shows

Figure 1.   Outline of the experimental approach used for the functional diversity array.

that the approach is sensitive enough to detect populations at naturally occurring levels. The use of phosphor-imaging screens, to which the entire array is exposed, allows ^{14}C bound to each spot to be quantified. We have experimentally determined the detection limit for ^{14}C on these screens to be 0.1 DPM for a spot 150 μm in diameter, which consists of about 10^8 oligonucleotide probes. Assuming that the rRNA molecules fragment to about 300 bp and that only 1% of the oligonucleotide probes will be bound to templates after hybridization, the detection limit is between 10^2 and 10^3 cells. This estimate is based on a cellular rRNA content between 1,000 and 10,000 molecules, which is in the range of slow- and fast-growing cells, respectively. Using this procedure, we have been able to specifically detect and differentiate rRNA from sulfate-reducing strains grown on ^{14}C-labeled lactate (Klepac and Polz, unpubl. data).

For the array to represent the actual microbial populations responsible for metabolism of a specific compound, realistic substrate concentrations must be used in the incubations to avoid introducing a major bias in community structure due to selective growth of specific populations. In coastal waters, which we are using as a model ecosystem, we found that even low additions of ^{14}C-labeled organic substrates (representing <3% of the total organic carbon) produced highly radiolabeled rRNA after 7 h incubation at *in situ* conditions. This incubation time is similar to the average

generation time for the entire bacterial community (9 h), and both the uptake rate and the growth yield on the labeled substrates were linear during the incubation, suggesting that there were no major shifts in the microbial community. These tests also showed that the proportion of labeled C allocated to rRNA was strongly dependent on the quality of the substrate (*e.g.*, 12%–19% for adenine, 1.1%–1.3% for acetate). In addition, tests with exponentially growing bacteria in pure culture showed that a constant fraction of the total cellular ^{14}C (average 8% for *Vibrio cholera* and 17% for *E. coli*) could be recovered in rRNA after about one cell doubling. Thus, the major advantages of rRNA detection are the linearity of the labeling process and the possible limitation to few cell doublings, which ensure that community structure will be only minimally biased.

The third question, whether rRNA diversity can be ascertained with realistic effort. requires a reexamination of the PCR-based approach. We have recently presented the hypothesis that a simple, PCR-induced artifact may lead to severe overestimation of diversity of rRNA genes (Thompson *et al.*, 2002). During the co-amplification of homologous templates with universal primers, a significant fraction (up to 50%) of products may be present as heteroduplices. These were increasingly prevalent as template diversity increased or primer availability became limiting (Thompson *et al.*, 2002). After cloning, heteroduplex molecules may

become subject to mismatch repair by the *E. coli* MutHLS system. This can theoretically lead to independent repair of each mismatched position since the repair system, *in vivo*, is directed by hemimethylation (Modrich, 1987), which is absent in PCR products. A model exploring the effects of heteroduplex repair demonstrated that the undirected repair process might be responsible for large overestimation of rRNA diversity (Thompson *et al.*, 2002). For example, a simple system of 2 sequences with 3 shared mismatched positions can result in 8 sequence permutations; for 4 sequences with 10 shared mismatched positions, the number increases to 6136 (Thompson *et al.*, 2002). Although this is a dramatic example, the potential contribution of heteroduplex repair to sequence diversity can easily be avoided by "reconditioning PCR," a low-cycle-number reamplification of a 10-fold diluted, mixed-template PCR product.

Although the exact contribution of heteroduplex repair to diversity estimates is still being analyzed in our laboratory, we have used the modified amplification protocol (reconditioning PCR) to estimate bacterial diversity in a coastal bacterial community. We generated a clone library from amplified 23S rRNA genes, then assessed sequence diversity in the library by a combination of rarefaction analysis and Chao-1 estimators, which are based on capture-recapture statistics (Hughes *et al.*, 2001). The results demonstrated that diversity was relatively moderate, with the number of coexisting sequence types remaining in the low 100s (Acinas, Hunt, Bertilsson, and Polz, unpubl. data). This ongoing analysis is currently complemented with rarefaction of a 16S rRNA gene library derived from the same sample, demonstrating that it may indeed be possible to represent entire communities with reasonable effort on DNA microarrays.

Acknowledgments

Literature Cited

Béja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrignes, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2: 516–529.

Bertilsson, S. A., and M. F. Polz. 2001. Application of a diversity array to study specific substrate utilization in individual populations of heterotrophic bacteria. P. 503 in *Abstracts*, American Society for Microbiology, 101st General Meeting, Orlando, FL.

Boschker, H. T. S., S. C. Nold, P. Wellsbury, D. Bos, W. de Graaf, R. Pel, R. J. Parkes, and T. E. Cappenberg. 1998. Direct linking of microbial populations to specific biogeochemical processes by ^{13}C-labeling of biomarkers. *Nature* 392: 801–805.

Cho, J.-C., and J. M. Tiedje. 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* 67: 3677–3682.

Cottrell, M. T., and D. L. Kirchman. 2000. Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* 66: 1692–1697.

Giovannoni, S., and M. Rappé. 2000. Evolution, diversity, and molecular ecology of marine prokaryotes. Pp. 47–84 in *Microbial Ecology of the Oceans*, D. L. Kirchman, ed. Wiley-Liss, New York.

Gray, N. D., and I. M. Head. 2001. Linking genetic identity and function in communities of uncultured bacteria. *Environ. Microbiol.* 3: 481–492.

Head, I. M., J. R. Saunders, and R. W. Pickup. 1998. Microbial evolution, diversity, and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms. *Microb. Ecol.* 35: 1–21.

Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* 180: 4765–4774.

Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67: 4399–4406.

Koizumi, Y., J. J. Kelly, T. Nakagawa, H. Urakawa, S. El-Fantroussi, S. Al-Muzaini, M. Fukui, Y. Urushigawa, and D. A. Stahl. 2002. Parallel characterization of anaerobic toluene- and ethylbenzene-degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridization, and DNA microarray technology. *Appl. Environ. Microbiol.* 68: 3215–3225.

Lee, N., P. H. Nielsen, K. H. Andreasen, S. Juretschko, J. L. Nielsen, K.-H. Schleifer, and M. Wagner. 1999. Combination of fluorescent in situ hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* 65: 1289–1297.

Modrich, P. 1987. DNA mismatch correction. *Annu. Rev. Biochem.* 56: 435–466.

Ouverney, C. C., and J. A. Fuhrman. 1999. Combined microautoradiography—16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ. *Appl. Environ. Microbiol.* 65: 1746–1752.

Radajewski, S., P. Ineson, N. R. Parekh, and J. C. Murrell. 2000. Stable-isotope probing as a tool in microbial ecology. *Nature* 403: 646–649.

Schmidt, T. M., E. F. DeLong, and N. R. Pace. 1991. Analysis of marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* 173: 4371–4378.

Small, J., D. R. Call, F. J. Brockman, T. M. Straub, and D. P. Chandler. 2001. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. *Appl. Environ. Microbiol.* 67: 4708–4716.

Stokes, H. W., A. J. Holmes, B. S. Nield, M. P. Holley, K. M. H. Nevalainen, B. C. Mabbutt, and M. R. Gillings. 2001. Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl. Environ. Microbiol.* 67: 5240–5246.

Thompson, J. R., L. A. Marcelino, and M. F. Polz. 2002. Heteroduplexes in mixed-template amplifications: formation, consequences and elimination by 'reconditioning PCR.' *Nucleic Acids Res.* 30: 2083–2088.

Torsvik, V., L. Ovreas, and T. F. Thingstad. 2002. Prokaryotic diversity: magnitude, dynamics, and controlling factors. *Science* 296: 1064–1066.